# Computer-Aided Tuberculosis Detection from Chest X-Ray Images with Convolutional Neural Networks

**Lucas Gabriel Coimbra Evangelista, Elloá B. Guedes**

[1]Grupo de Pesquisas em Sistemas Inteligentes
Escola Superior de Tecnologia
Universidade do Estado do Amazonas
Av. Darcy Vargas, 1200 – Manaus – Amazonas

{lgce.eng, ebgcosta}@uea.edu.br

***Abstract.*** *Diagnosing Tuberculosis is crucial for proper treatment since it is one of the top 10 causes of deaths worldwide. Considering a computer-aided approach based on intelligent pattern recognition on chest X-ray with Convolutional Neural Networks, this work presents the proposition, training and test results of* $9$ *different architectures to address this task as well as two ensembles. The highest performance verified reaches accuracy of* $88.76\%$*, surpassing human experts on similar data as previously reported by literature. The experimental data used comes from public medical datasets and comprise real-world examples from patients with different ages and physical characteristics, what favours reproducibility and application in practical scenarios.*

## 1. Introduction

*Tuberculosis* (TB) is chronic infectious disease, that most often affects the lungs, caused by *Mycobacterium tuberculosis*. It is airborne spread from person to person, and about one-third of the world's population has its latent form, which means they have been infected by TB but are not (yet) ill and cannot transmit it [Jamison et al. 2006]. According to World Health Organization (WHO), TB is one of the top $10$ causes of death worldwide, surpassing HIV[1] and malaria [WHO 2016].

Brazil, in particular, is among the countries with the highest number of TB cases in the world and, since 2003, this disease has been considered as a priority in the political agenda of the Brazilian Ministry of Health. Although it is a disease with diagnosis and treatment performed universally and free of charge by the Unified Health System, in Brazil there are still many practical problems that result in approximately $69.000$ new cases and $4.500$ deaths each year caused by TB. Although still high from a global perspective, the number of TB cases in Brasil has been decreasing due to the efforts of National Tuberculosis Control Program (NTCP), developed in partnership with states, cities and civil society [Brazilian Ministry of Health 2016a].

According to the data available, it is possible to identify some groups more susceptible to TB in Brazil due to social vulnerability, such as: $(i)$ prison population, where the TB incidence is 28 times greater than average; $(ii)$ homeless people; $(iii)$ TB-HIV coinfected pacients, with incidence of $3.6$ per $100$ thousand inhabitants; $(iv)$ indigenous population, with $3$ times the average incidence; and $(v)$ health professionals. The black population in Brazil also suffers a $2.2$ higher incidence of TB than the general population [Brazilian Ministry of Health 2016b].

Despite the several challenges regarding this disease, ending the TB epidemic by $2030$ is among the health targets of the Sustainable Development Goals

[1] Human Immunodeficiency Virus

[United Nations 2015]. In order to contribute with this goal, new treatments and diagnosis strategies need to be developed.

When considering computer-aided approaches to support TB diagnosis, several challenges must be taken into account. The first of them is to consider anatomical shape variations, for example, in heart dimensions, costodiaphragmatic recess, rib cage and clavicle bones. X-ray imaging inhomogeinities are very common, specially considering the shortage of radiological infrastructure and radiologists in some areas, or even the technological differences between X-ray equipments. Lastly, the existence of other patologies besides TB, such as pneumonia, may result in variations in lung appearance [Jaeger et al. 2014a].

Considering the massive influx of multimodality data available nowadays, Deep Learning (DL) approaches are forging outstanding results on computer aided diagnosis [Ravì et al. 2017]. Positive results in Radiology, in particular, are helping in many tasks such as in the identification of pulmonary nodules, classification of breast density on mammograms, among others [McBee et al. 2018, Yasaka et al. 2018]. The case for TB detection in chest X-ray images has also been addressed with Convolutional Neural Networks (CNNs) by using pre-trained models and augmented data and achieving impressive results [Lakhani and Sundaram 2017]. However, if one takes into account the computational cost, the misdiagnosis risks some data augmentation operations can introduce and the reprodutibility of the results that rely on private datasets, further research on the topic is still needed.

In this perspective, this work aims at using CNNs on several realistic anteroposterior X-ray images from public domain datasets aiming at distinguish TB cases from healthy ones, contributing to the computer-aided diagnosis of such disease. The results obtained consider $9$ different CNNs from two high-level architectures with no data augmentation nor transfer learning. The higher accuracy verified was equal to $88.76\%$ in an individual learner with a single convolutional layer using $32$ kernels. These results surpass human specialists and may indicate that a profound abstract feature representation is not absolutely determinant for the performance on this task. Two ensembles built from the individual learners previously trained were also proposed and tested, with comparable accuracy metrics.

In order to present such results, this paper is organized as follows: some efforts in the literature that address computer-aided TB diagnosis are presented in Section 2; the experimental data collection, preparation and descriptive characteristics are detailed in Section 3; the methodology adopted to conduct this work is introduced in Section 4; results obtained are presented and discussed in Section 5; lastly, final remarks and future work are shown in Section 6.

## 2. Related Work

As in many other domains, the first solutions for computer-aided TB diagnosis comprised expert systems. Expert systems are a branch of applied Artificial Intelligence whose basic idea is to transfer human knowledge on a specific task to a computer [Liao 2005]. In this perspective, these works considered symptons and their intensity, such as fever, abdominal pain, skin lesions, hemoptysis, and others, in a rule-based system generally used for premedical care [Imianvan and Obi 2011, Agah 2013]. This approach, however, is strongly dependent on human knowledge and intervention, clinical findings and also on diagnostics tests.

More autonomous approaches considered by literature take into account that findings on chest X-ray images, an almost inexpensive exam, are associated with manifestations of active TB, such as: cavity formation, enlargement of airways, miliary pattern, lymph node enlargement, etc. This way, many approaches based on automatic pattern recognition were developed aiming tasks of lung segmentation, bone supression, lung boundary detection and feature extraction and classification. Although the results obtained so far, it is a difficult task to quantify the progress because some datasets are not publicly available, the X-ray images conditions may differ, among other difficulties [Jaeger et al. 2013].

A solution proposed by Jaeger et al. aims at automatic TB screening from chest X-rays in which lung detection and feature extraction are performed before classification with Machine Learning models (support vector machines, artificial neural networks, logistic regression and decision trees) [Jaeger et al. 2014b]. They use chest X-ray images from two public datasets [Jaeger et al. 2014a], and they also propose a comparison amongst the performance of the proposed model and of radiologists. The results observed, with accuracies of $78.3\%$ and $84\%$ for each dataset, respectively, were superior than human performance and is being under use in remote areas of Kenya.

Recently, with the advances in hardware and software as well as with the crescent amount data available, computers have allowed to perform an increasing number of complex tasks [Goodfellow et al. 2016]. Deep Learning (DL) algorithms are an instance of that, with a growing number of applications in Health Informatics [Ravì et al. 2017], including the field of Radiology such as in lesion or disease detection, classification and diagnosis, segmentation, and quantification [McBee et al. 2018].

The work of Lakhani and Sundaram is one of the earliest to address TB detection with Deep Learning using Convolutional Neural Networks [Lakhani and Sundaram 2017]. The authors used four datasets with labelled examples and adopted canonical convolutional neural networks architectures, such as AlexNet and GoogLeNet, and also tested whether the results would improve with and without data augmentation and transfer learning. The best results were verified with pre-trained weights and data augmentation on an ensemble model with two CNNs, with AUC of $0.99$. These results are remarkable but the authors emphashize that they do not replace human radiologic interpretation beyond that of TB .

If the number of calculations are taken into account, the work of Lakhani and Sundaram can be considered as having a high computational cost because of the number of operations performed due to the depth of the models under use [Lakhani and Sundaram 2017]. Another aspect to consider is that they use images from four datasets in which one of them, obtained from the Thomas Jefferson University, is not publicly available, compromising the reprodutibility of the results and making it difficult to compare with prior work on the topic.

## 3. Experimental Data

In order to comprise a realistic experimental dataset to train and test convolutional neural networks, posteroanterior chest radiographs were considered. This kind of radiography is widely adopted for diagnosis purposes and consists in a projection radiograph of the patient's chest. These images were gathered from the following public health images databases:

1. **JSRT Database**. Created and mantained by the Japanese Society of Radiological

Technology (JSRT), this dataset contains 247 images from which 154 contain pulmonary nodules and 93 are clean. All images have high-resolution, 2048 × 2048 pixels in grayscale, and metadata containing patient gender, diagnosis, among other information [Shiraishi et al. 2000];

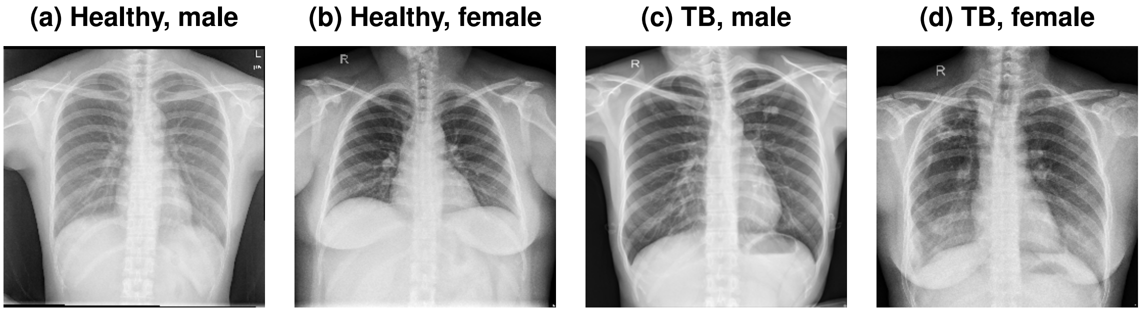
2. **Montgomery County X-ray Set**. X-ray images in this data set have been acquired from the tuberculosis control program of the Department of Health and Human Services of Montgomery County, Maryland, USA. This set contains 138 posteroanterior X-rays, of which 80 are normal and 58 are abnormal with different levels of tuberculosis manifestations [Jaeger et al. 2014a];
3. **Shenzhen Hospital X-ray Set**. X-ray images in this data set have been collected by Shenzhen No.3 Hospital in China as part of the routine care. There are 326 normal and 336 abnormal images showing various manifestations of tuberculosis, contemplating 21 pediatric patients [Jaeger et al. 2014a].

After data gathering phase, the images and their available metadata were inspected. This processed revealed that not all pulmonary nodules documented in JSRT database were originary from TB. In order to avoid misdiagnosis, it was decided that examples in this category would be discarded.

The next step considered image standardization through rescaling because images from both Montgomery County and Shenzhen datasets had different sizes. For diagnosis purposes the literature suggests dimensions from 128 × 128 up to 1024 × 1024 pixels. In this work, we considered 256 × 256 pixels due to processing time required by the Machine Learning models considered and to avoid overfitting, upon providing a possible considerable ammount of irrelevant features.

As a result, the available experimental data to train and test the models contains 893 samples, from which 394 are suggestive for TB according to experts and the remaining part is composed of normal cases. As it can be seen, this database is both labelled and unbalanced, since all examples have an associate label (TB or Healthy) and the amount of examples per labels is not evenly distributed. The dataset consolidated for this work is realistic for computer-aided diagnosis purposes of TB with Machine Learning. It comprises several manifestations of TB around the world, from male and female subjects with different ages and distinct physical characteristics, as illustrated in Figure 1. Moreover, the labels were accredited from experts and images were acquired with typical X-ray devices.

**Figure 1: Four distinct samples of chest X-rays in the dataset consolidated.**



(a) Healthy, male

(b) Healthy, female

(c) TB, male

(d) TB, female

## 4. Materials and Methods

This section aims at presenting the material and methods considered for this work. In the first part, the procedures regarding the use of the available experimental data are presented. After that, the proposal of CNNs models and their parameters are introduced. Then, the performance metrics to evaluate the models are shown. At last, an ensemble approach to combine individual learners is also considered.

The experimental data available, described in the previous section, will be randomized and later divided according to a hold-out validation method: training data, corresponding to $70\%$ of examples, will be used to provide experience regarding TB and healthy examples; test data, with the remaining $30\%$ of examples, will be used to evaluate the CNNs in the classification task. In order to accurate model training and to estimate model properties, $10\%$ of training data will be reserved for validation purposes [Brink et al. 2017].

Recalling the role of dataset size in models' performance, a typical approach adopted is *data augmentation* in which more training data is generated from the existing samples via a number of random transformations (rotations, horizontal flip, random shifts, etc.). In many situations, data augmentation yelds believable-looking images and helps the models to generalize better, helping in avoiding overfitting [Chollet 2017].

Althought it is considered to improve the generalization capabilities of machine learning models specially in Computer Vision tasks, data augmentation will not be considered in the scope of this work because of the risks of misdiagnosis it can introduce. For example, a zooming inside a medical image may discard relevant information that lies in another part of it. Horizontal flips, for instance, violate the assumption of horizontal asymmetry which may counfound the *situs inversus* condition.

The machine learning models considered in this work to address TB diagnosis from chest images are the *Convolutional Neural Networks* (CNNs). These models are analogous to Artificial Neural Networks but each unit in a layer is a high-dimensional filter which is convolved with the input of that layer. These filters incorporate spatial context by having a similar (but smaller) spatial shape as the input media, and use parameter sharing to significantly reduce the number of learnable variables. CNNs are a prime example of deep learning methods currently state-of-the-art for Computer Vision [Khan et al. 2018].

Considering the different kinds of layers and the ways to dispose them, different architectures of CNNs can be proposed to address the TB pattern recognition in the images from the dataset. Two high-level architectures were considered, where the first is:

$$\text{Input Layer} \Rightarrow (\text{Convolution} \rightarrow \text{Max-Pooling})^k \Rightarrow \text{Dense Layer} \Rightarrow \text{Softmax},$$

where $k = 1, \ldots, 5$ indicates the number of repetitions of the associated block of layers. In particular, when $k = 1$ we have the most basic filtering approach to feature extraction. The Convolutional Layer has $32$ units using $3 \times 3$ kernels and stride equal to $1$. In all cases, the Dense Layer has $128$ neurons and the Output Layer has a Softmax function in accordance with the binary classification task considered. According to this first high level architecture and selected parameters, five different CNNs were proposed, one for each $k$ value.

The second high-level architecture considers two sequential convolutional operations as follows, where $j, \ell = 1, 2$:

$$\text{Input Layer} \Rightarrow (\text{Convolution} \rightarrow \text{Convolution} \rightarrow \text{Max-Pooling})^j \Rightarrow \text{Dense Layer}^\ell \Rightarrow \text{Softmax}.$$

Resuming this strategy of model proposal, nine different CNNs will be considered for the task. No such architectures were found with pre-trained data, so transfer learning could not be applied. Considering that all CNNs proposed will undergo a full training procedure with early stopping criteria, we used $100$ epochs, learning rate $\eta = 10^{-3}$ which was experimentally obtained and binary cross-entropy as loss measure.

In the binary classification task considered in this work with unbalanced classes, two performance metrics will be used to evalute the proposed CNNs on the test set: accuracy and micro F-Score. The former denotes the proportion of correct classification amongst all classifications, providing an overview of the model quality. The later, in turn, is particularly more specific for binary tasks since it quantifies the harmonic average between precision and recall per class alongside the prevalence proportion of each class [Kubat 2015]. Thus, the CNNs will be ranked according to the micro F-Score in the test set and their quality to address the original problem will be evaluated and discussed.

Besides using individual CNNs, two homogeneous ensembles will also be proposed: one ensemble using all previously mentioned CNNs and the other using the Top-3 CNNs ranked. These ensembles will combine the individuals outputs upon majority voting. Ensemble methods use multiple learners of the same problem and combine them for use, being significantly more accurate than a single learner in many real-world tasks [Zhou 2012]. The ensembles will also be evaluated according to the performance metrics under consideration.

Regarding implementation, the Python programming language [Python 2018], the open-source frameworks Sci-Kit Learn [Pedregosa et al. 2011] and Keras [Keras 2018] and the Jupyter Notebook[2] interactive development environment were adopted.

## 5. Results and Discussion

Upon conducting the steps presented in the previous sections, the following results were obtained. Firstly, the CNNs proposed were implemented and trained by following the procedures described. Data presented in Table 1 depicts the details of such CNNs in means of trainable parameters and number of epochs until early stopping, when occurred.

**Table 1: CNNs proposed and their respective trainable parameters and learning epochs.**

| High-Level Architecture | k | j | $\ell$ | Trainable Parameters | Epochs |
|---|---|---|---|---|---|
| 1 | 1 | - | - | 66,065,666 | 80 |
| 1 | 2 | - | - | 15,755,554 | 100 |
| 1 | 3 | - | - | 3,706,168 | 100 |
| 1 | 4 | - | - | 831,842 | 100 |
| 1 | 5 | - | - | 185,730 | 100 |
| 2 | - | 1 | 1 | 65,038,626 | 99 |
| 2 | - | 1 | 2 | 65,055,138 | 96 |
| 2 | - | 2 | 1 | 15,270,242 | 100 |
| 2 | - | 2 | 2 | 15,286,754 | 96 |

Performance metrics obtained on test data are synthesized in Table 2. As it can be seen, $8$ out of $9$ CNNs proposed have accuracy higher than $80\%$. The columns $TP$ (true positive), $FP$ (false positive), $FN$ (false negative) and $TN$ (true negative) accounts the values of hits and misses in this binary task with $267$ test examples.

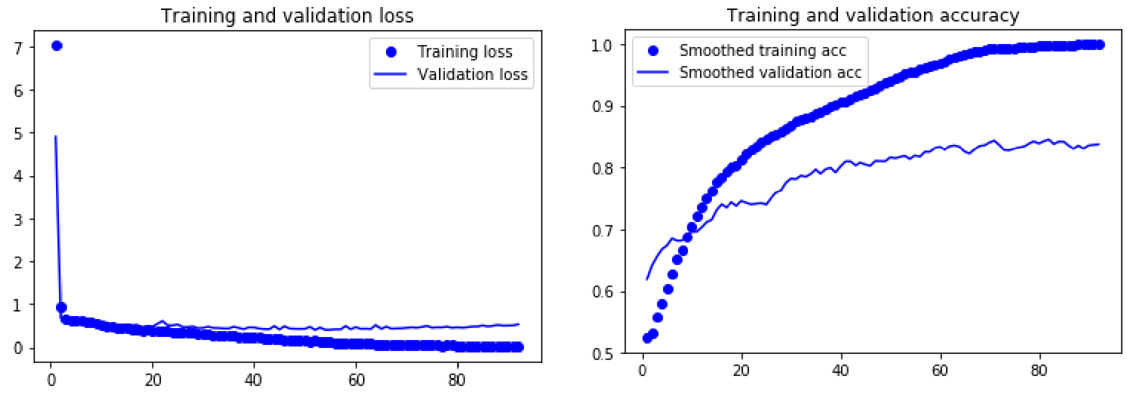
[2] http://jupyter.org/

Table 2: Performance metrics on test data.

| High-Level Architecture | k | j | $\ell$ | TP | FP | FN | TN | Accuracy | Micro F-Score |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | - | - | 98 | 14 | 16 | 139 | 88.76% | 0.8876 |
| 1 | 2 | - | - | 104 | 27 | 10 | 126 | 86.14% | 0.8614 |
| 1 | 3 | - | - | 104 | 33 | 10 | 120 | 83.89% | 0.8389 |
| 1 | 4 | - | - | 87 | 26 | 27 | 127 | 80.14% | 0.8014 |
| 1 | 5 | - | - | 102 | 57 | 12 | 96 | 74.15% | 0.7415 |
| 2 | - | 1 | 1 | 88 | 11 | 26 | 142 | 86.14% | 0.8614 |
| 2 | - | 1 | 2 | 86 | 14 | 28 | 139 | 84.26% | 0.8426 |
| 2 | - | 2 | 1 | 91 | 20 | 23 | 133 | 83.89% | 0.8389 |
| 2 | - | 2 | 2 | 91 | 22 | 23 | 131 | 83.14% | 0.8314 |

The CNN obtained from the first high-level architecture with $k = 1$ had best performance amongst the models proposed, with training and validation losses and accuracies detailed in Figure 2. This CNN also has the highest quantity of trainable parameters, but does have sequential convolutional layers that perform high sequential feature map extraction. It may indicate that relevant TB characteristics can be extracted in few convolutional levels.

Figure 2: Binary cross-entropy loss and accuracy per epoch collected in training and validation phases of the CNN with best observed performance.



(a) Loss

(b) Accuracy

The performance of the solution proposed by Jaeger et al. [Jaeger et al. 2014b] with lung segmentation and feature extraction prior TB classification with Machine Learning models has accuracy of 78.3% and 84% for Montgomery Count and Shenzhen Hospital X-ray sets, respectively, surpassing human specialists performance. Considering the performance observed, 4 CNNs proposed also overcome this result, therefore exceeding also human experts. It should be noticed that only the raw chest images are considered, no pre-processing is made for purposes of feature extraction.

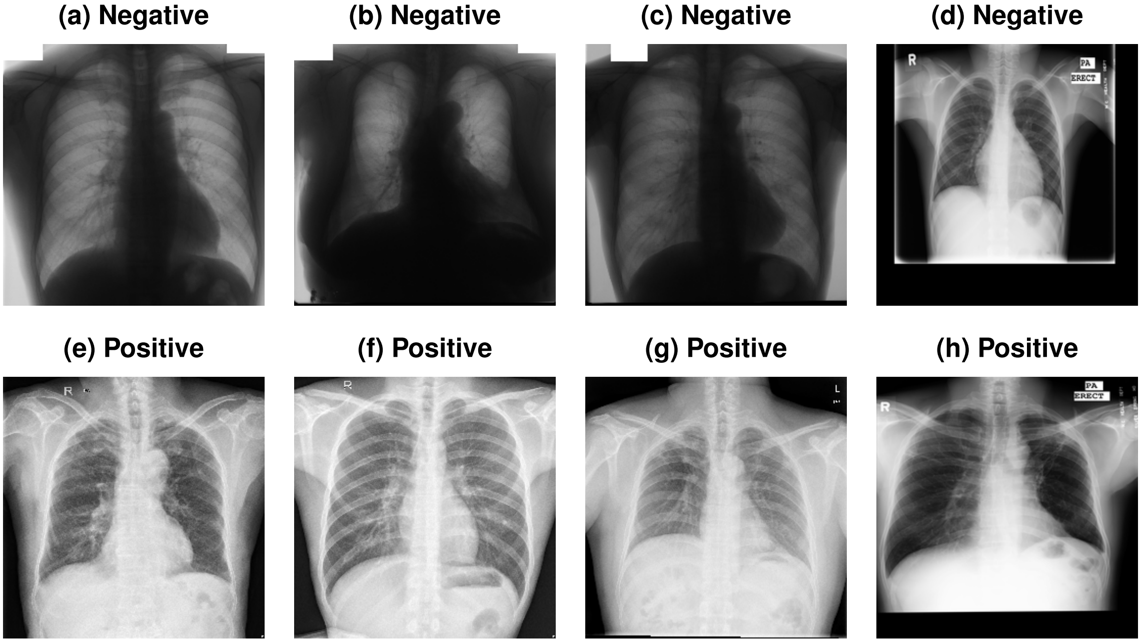
Upon combining these trained models under two homogeneous ensembles, the results obtained on test set are shown in Table 3. The Top-3 CNNs ensemble has superior performance than the All CNNs ensemble, but with less operations. It was also expected that the ensemble would have superior performance than the individual learners, but it may not occurred because the learners are not distinct enough to capture non-overlapping features.

**Table 3: Ensembles performance on test data.**

| Ensemble | TP | FP | FN | TN | Accuracy | Micro F-Score |
|---|---|---|---|---|---|---|
| All CNNs | 91 | 22 | 23 | 131 | 83.14% | 0.8651 |
| Top-3 CNNs | 88 | 11 | 26 | 142 | 86.14% | 0.8839 |

Taking a closer look in the Top-3 CNNs ensemble decisions, some particular cases emerge. The 3 CNNs agreed correctly in 79.77% of the correct predictions performed wheter positive or negative for TB. However, in 19 particular cases, all three CNNs had unanimous wrong votes. Taking a closer look, 11 cases were of healthy examples mistaken with TB and the rest were labelled as healthy while being TB positive. Some of these cases are shown in Figure 3.

**Figure 3: Particular cases of wrong ensemble decision.**



(a) Negative (b) Negative (c) Negative (d) Negative

(e) Positive (f) Positive (g) Positive (h) Positive

As it can be seen, in some of these cases the chest X-ray images are not centralized, have inverted color pattern and even comprehend one case of pediatric subject. It represented challenges for both individual and ensemble learners. Regarding these images, most of these cases (68%) were from Shenzhen Hospital X-ray Set. So, although differences in X-ray equipments technology, it is important to avoid inhomogeneities to favour automatic approaches for pattern recognition.

## 6. Final Remarks

In this work we addressed TB detection in chest X-ray images with convolutional neural networks. Nine different CNNs were proposed as well as two ensembles. By using realistic medical images from two public available datasets accredited by radiologists, the results obtained indicate an accuracy of $88.76\%$, surpassing techniques with feature extraction and human expertise on this problem. The CNNs were proposed according to two high-level architectures and in their training no data augmentation was used. The best model for this task is an individual CNN whose computational cost is lower than the corresponding ensemble.

As a contribution to further research on computer-aided TB diagnosis, the proposed models with pre-trained weights on these datasets can be obtained in (GitHub Repo link). Future work aim at parameters fine-tuning and also at training and testing other CNN architectures. Other perspective considered in the next steps comprise an analysis of performance of such models in pulmonary TB cases prevalent in Brazil.

## Acknowledgements

The authors thank the financial support provided by FAPEAM through PAIC/UEA 2017-2018 and Research Grant PROTI PESQUISA 11/2017.

## References

Agah, A. (2013). *Medical Applications of Artificial Intelligence*. CRC Press, United States.

Brazilian Ministry of Health (2016a). Brasil Livre da Tuberculose. Available at `http://portalarquivos.saude.gov.br/images/pdf/2017/fevereiro/24/Plano-Nacional-Tuberculose.pdf`.

Brazilian Ministry of Health (2016b). Panorama da Tuberculose no Brasil. Available at `http://bvsms.saude.gov.br/bvs/publicacoes/panorama%20tuberculose%20brasil_2014.pdf`.

Brink, H., Richards, J. W., and Fetherolf, M. (2017). *Real-World Machine Learning*. Manning Publications, United States.

Chollet, F. (2017). *Deep Learning with Python*. Manning Publications, Shelter Island, New York, 1 edition.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.

Imianvan, A. A. and Obi, J. (2011). Fuzzy cluster means expert system for the diagnosis of tuberculosis. *Global Journal of Computer Science & Technology*, 11(6):41–48.

Jaeger, S., Candemir, S., Antani, S., Yi-Xiang, Wang, J., Lu, P.-X., , and Thoma, G. (2014a). Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quantitative Imaging in Medicine and Surgery*, 6(4):475–477.

Jaeger, S., Karargyris, A., Candemir, S., Folio, L., Siegelman, J., Callaghan, F., Xue, Z., Palaniappan, K., Singh, R. K., and Antani, S. (2014b). Automatic tuberculosis screening using chest radiographs. *IEEE Transactions on Medical Imaging*, 33(2):233–245.

Jaeger, S., Karargyris, A., Candemir, S., Siegelman, J., Folio, L., Antani, S., and Thoma, G. (2013). Automatic screening for tuberculosis in chest radiographs: a survey. *Quantitative Imaging in Medicine and Surgery*, 3(2):89–99.

Jamison, D. T., Breman, J. G., Measham, A. R., Alleyne, G., Claeson, M., Evans, D. B., Jha, P., Mills, A., and Musgrove, P. (2006). *Disease Control Priorities in Developing Countries*. World Bank Publications, 2th edition.

Keras (2018). Keras: The python deep learning library. Available at http://keras.io/. Accessed in August 14, 2018.

Khan, S., Rahmani, H., Shah, S. A. A., and Bennamoun, M. (2018). *A Guide to Convolutional Neural Networks for Computer Vision*. Morgan and Claypool.

Kubat, M. (2015). *An Introduction to Machine Learning*. Springer, United States.

Lakhani, P. and Sundaram, B. (2017). Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, 284(2):1–9.

Liao, S.-H. (2005). Expert system methodologies and applications — a decade review from 1995 to 2004. *Expert systems with applications*, 28(1):93–103.

McBee, M. P., Awan, O. A., Colucci, A. T., Ghobadi, C. W., Kadom, N., Kansagra, A. P., Tridandapani, S., and Auffermann, W. F. (2018). Deep learning in radiology. *Academic Radiology*, March 30rd, 2018:1–9. In Press.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Python (2018). Python programming language. Available at http://www.python.org. Accessed in August 14, 2018.

Ravì, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., and Yang, G. Z. (2017). Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, 21(1):4–21.

Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K., Matsui, M., Fujita, H., Kodera, Y., and Doi, K. (2000). Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *American Journal of Roentgenology*, (174):71–74.

United Nations (2015). Sustainable development goals. Available at https://www.un.org/sustainabledevelopment/sustainable-development-goals/.

WHO (2016). Global Tuberculosis Report 2016. Available at http://www.who.int/tb/publications/global_report/en/.

Yasaka, K., Akai, H., Kunimatsu, A., Kiryu, S., and Abe, O. (2018). Deep learning with convolutional neural network in radiology. *Japanese Journal of Radiology*, 26(257):1–16.

Zhou, Z.-H. (2012). *Ensemble Methods – Foundations and Algorithms*. CRC Press, United States.